A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 235, 2010

Dorota Pruska[*]

ROBUSTNESS OF THE DEEPEST REGRESSION METHOD WITH RESPECT TO OUTLIERS FOR SELECTED SAMPLING SCHEMES

Abstract. The deepest regression method is a method of estimation of regression parameters which is robust with respect to outliers in samples drawn from the population.

In this paper, the simulation analysis of robustness of the deepest regression method for outliers for selected sampling schemes was conducted. On the basis of the Monte Carlo experiments the characteristics of distribution of regression parameter estimates obtained by deepest regression method for the samples containing outliers were determined. The results were compared with the results of analogous experiments conducted with the usage of the least square method (LSM), where the outliers were removed from the samples prior to estimation. The deepest regression method turned out to be robust with respect to the outliers in all considered cases, and the outcomes were similar to the results obtained with LSM after removing outliers from samples.

Key words: the deepest regression method, outliers, sampling schemes

I. ROBUSTNESS OF THE DEEPEST REGRESSION METHOD WITH RESPECT TO OUTLIERS

The deepest regression method (DRM) is a method of estimation regression parameters in such a way, that the maximal regression depth characterizes the obtained model (see Rousseeuw, Hubert (1999)). In applications of DRM the algorithm MEDSWEEP can be used to determine the estimates of parameters of the following linear model (see Van Aelst et al. (2000)):

$$y = \theta_1 x_1 + ... + \theta_{p-1} x_{p-1} + \theta_p + \varepsilon\,, \tag{1}$$

describing dependence between variables Y and X, whose realizations are respectively: y_i and $(x_{i1}, ..., x_{i,p-1})$, i=1,..., n in considered data set $Z_n = \{(x_{i1}, ..., x_{i,p-1}, y_i)$; $i = 1,..., n\} \subset R^p$ and θ is a vector of model parameters, $\theta=[\theta_1,..., \theta_p]^T$.

The deepest regression method is robust with respect to outliers in data set, if their fraction does not exceed the breakdown value ε_n^*. The breakdown value

[*] MA, PhD student in Chair of Statistical Methods, University of Łódź.

ε_n^* of an estimator T_n is defined as the smallest fraction of outliers which, added to the n-element data set Z_n, make the estimator unrobust (see Van Aelst et al. (2000)). Let Z_{n+m} be the data set obtained by adding m outliers to Z_n. The breakdown value is of the form:

$$\varepsilon_n^*(T_n, Z_n) = \min\left\{\frac{m}{m+n}; \sup_{Z_{n+m}} \left\| T_{n+m}(Z_{n+m}) - T_n(Z_n) \right\| = \infty \right\}. \tag{2}$$

Let $Z_n = \{(\mathbf{x}_i^{\mathrm{T}}, y_i);\ \mathbf{x}_i^{\mathrm{T}} \in R^{p-1},\ y_i \in R,\ i = 1,...,n\}$ be a sample from the population. In case of DRM-estimator T_r^*, for $p \geq 2$, the breakdown value has the following property (see Van Aelst et al. (2000)):

$$\varepsilon_n^*(T_r^*, Z_n) \underset{n\to\infty}{\overset{a.s.}{\to}} \frac{1}{3}, \tag{3}$$

where *a.s.* denotes 'almost surely'.

II. MONTE CARLO EXPERIMENTS FOR SELECTED SAMPLING SCHEMES

The aim of this paper is to analyze the robustness of the deepest regression method for selected sampling schemes using the simulation experiments and comparing the results with the outcomes of least square method obtained for the same samples, but after removing the outliers. Three groups of Monte Carlo experiments were conducted, in which some procedures from the algorithm MEDSWEEP and some procedures presented by Brandt (1999) were used.

The first group of experiments, named *E1*, was conducted as follows:

1. Construction of the N-element population (N = 2000) described by the random variable (X_1, X_2, Y), where values of X_1, X_2 were generated independently from the chi-square distributions respectively: χ^2_5 and χ^2_{10}, and values of Y variable was constructed according to the formula:

$$y_i = 2x_{1i} + 3\,x_{2i} + 1 + \varepsilon_i \quad \text{for} \quad i = 1,\ldots, N, \tag{4}$$

   where ε_i is a random error which is normally distributed $N(0; 0{,}5)$.
2. Drawing with replacement and without replacement a n-element sample (n = 100, 200, 300, 400, 500, 1000).

3. Input of the outliers into the samples by multiplication the first k values of variable Y by 1,5 (three fractions of outliers k/n = 0,1; 0,2; 0,3 were considered).
4. DRM-estimation (for samples with outliers) and LSM-estimation (for samples from which outliers were removed) of the parameters of model:

$$y = a_1 x_1 + a_2 x_2 + a_3 + \varepsilon \tag{5}$$

5. Calculating relative error for a sample (RE_s) and relative error for a population (RE_{pop}) according to the formulas:
6.

$$RE_s = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

and

$$RE_{pop} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{7}$$

where $\hat{y}_i$ is the theoretical value of variable Y obtained on the basis of model (5).

7. The 10000 repetitions of stages 2-5 for each considered sample size, fraction of outliers and sampling scheme.
8. Calculating the mean (*MEAN*), the coefficient of variation (V) and the relative root mean square error (*RRMSE*) for 10000 estimates for each parameter of model (5), the mean of relative errors for sample ($\overline{RE}_s$) and the mean relative errors for population ($\overline{RE}_{pop}$) for each considered sample size, fraction of outliers and sampling scheme.
9. Calculation of efficiency growth (*EFG*) for estimator of each parameter of model (5) obtained while drawing samples without replacement instead of with replacement, given by the formula (see Zasępa (1972)):

$$EFG = \frac{D_0^2(\theta) - D^2(\theta)}{D^2(\theta)} \cdot 100\% \ , \tag{8}$$

where $D_0^2(\theta)$ is a variance of 10000 estimates of parameter θ for sample drawn with replacement, and $D^2(\theta)$ a variance of 10000 estimates of parameter θ for sample drawn without replacement.

The results of the experiments *E1* are presented in tables 1.–5. (statistical characteristics of estimates and values of *EFG*) and on the graph 1 (illustration of *EFG* for $k/n = 0{,}2$).

Table 1. Statistical characteristics of DRM-estimates of parameters of model (5) for the experiments *E1*, k/n=0,2 and samples drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	1,992	0,019	0,020	3,003	0,009	0,009	1,187	0,282	0,384	0,075	0,012
200	1,992	0,013	0,014	3,002	0,006	0,006	1,192	0,196	0,302	0,075	0,012
300	1,992	0,010	0,011	3,002	0,005	0,005	1,192	0,157	0,268	0,075	0,012
400	1,992	0,009	0,010	3,002	0,004	0,004	1,193	0,136	0,253	0,075	0,012
500	1,991	0,008	0,009	3,002	0,004	0,004	1,197	0,122	0,245	0,075	0,012
1000	1,991	0,006	0,007	3,001	0,003	0,003	1,203	0,089	0,230	0,075	0,012

Source: author's calculations.

Table 2. Statistical characteristics of DRM-estimates of parameters of model (5) for the experiments *E1*, k/n=0,2 and samples drawn without replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	1,992	0,019	0,019	3,003	0,009	0,009	1,188	0,277	0,379	0,075	0,012
200	1,991	0,013	0,013	3,003	0,006	0,006	1,192	0,185	0,292	0,075	0,012
300	1,991	0,010	0,011	3,002	0,005	0,005	1,194	0,150	0,264	0,075	0,012
400	1,991	0,008	0,009	3,002	0,004	0,004	1,195	0,129	0,249	0,075	0,012
500	1,991	0,007	0,008	3,002	0,004	0,004	1,196	0,113	0,238	0,075	0,012
1000	1,990	0,005	0,007	3,001	0,002	0,002	1,206	0,076	0,225	0,075	0,012

Source: author's calculations.

Table 3. Statistical characteristics of LSM-estimates of parameters of model (5) for the experiments *E1*, k/n=0,2 and samples drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	1,999	0,010	0,010	3,003	0,004	0,004	0,991	0,166	0,165	0,011	0,011
200	1,999	0,007	0,007	3,003	0,003	0,003	0,988	0,115	0,114	0,011	0,011
300	1,999	0,006	0,006	3,003	0,002	0,003	0,988	0,093	0,093	0,011	0,011
400	2,000	0,005	0,005	3,003	0,002	0,002	0,987	0,081	0,081	0,011	0,011
500	1,999	0,004	0,004	3,003	0,002	0,002	0,988	0,072	0,072	0,011	0,011
1000	2,000	0,003	0,003	3,003	0,001	0,002	0,988	0,051	0,052	0,011	0,011

Source: author's calculations.

Table 4. Statistical characteristics of LSM-estimates of parameters of model (5) for the experiments *E1*, *k/n*=0,2 and samples drawn without replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	1,999	0,010	0,010	3,003	0,004	0,004	0,989	0,161	0,160	0,011	0,011
200	1,999	0,006	0,006	3,003	0,003	0,003	0,989	0,109	0,109	0,011	0,011
300	1,999	0,005	0,005	3,003	0,002	0,002	0,988	0,088	0,087	0,011	0,011
400	2,000	0,004	0,004	3,003	0,002	0,002	0,987	0,074	0,074	0,011	0,011
500	2,000	0,004	0,004	3,003	0,002	0,002	0,988	0,064	0,065	0,011	0,011
1000	2,000	0,002	0,002	3,003	0,001	0,001	0,987	0,039	0,041	0,011	0,011

Source: author's calculations.

Table 5. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E1* and *k/n* = 0,1; 02; 03

Sample size	Values of *EFG* for parameters																	
	k/n=0,1						*k/n*=0,2						*k/n*=0,3					
	A1		A2		A3		A1		A2		A3		A1		A2		A3	
	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM
100	2,4	6,3	5,3	5,8	4,5	6,3	2,4	4,6	5,6	7,1	3,9	6,5	-2,9	4,5	-0,8	6,7	-0,9	6,4
200	10,8	10,1	14,2	9,7	14,9	11,0	11,4	10,1	8,4	8,5	12,3	10,2	6,3	8,0	5,8	8,4	7,4	9,3
300	14,6	17,0	11,6	15,4	13,2	15,6	13,0	14,8	6,6	13,3	9,1	12,4	8,3	12,6	8,0	11,5	10,3	12,3
400	18,9	21,2	19,8	21,4	20,8	25,6	13,1	18,2	11,7	18,8	11,6	21,0	11,5	16,7	9,4	16,2	10,8	17,7
500	22,4	26,0	28,8	28,6	28,2	27,4	19,2	20,3	17,9	29,2	16,9	24,2	11,4	17,4	13,4	20,7	13,8	20,8
1000	63,0	77,8	69,6	85,7	59,2	81,5	42,4	63,6	36,5	66,7	37,0	67,9	21,1	51,9	22,5	54,5	27,6	56,2

Source: author's calculations.

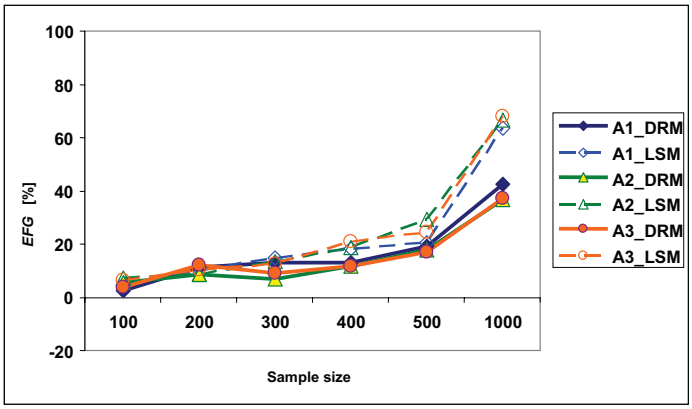

Fig. 1. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E1* and *k/n* = 0,2

Source: Table 5.

The group of experiments *E2*, is analogous to the previous one, and was conducted as follows:

1. Construction of the *N*-element (N = 2000) population described by the random variable (X_1, X_2, Y), where values of X_2 were generated independently from the chi-square distribution χ^2_5, $X_1 = X_2^2$ and Y variable was constructed according to the formula (4).
2. Drawing with replacement and without replacement a *n*-element sample (n = 100, 200, 300, 400, 500, 1000).
3. Input of the outliers into the samples by replacement the first k values of variable Y by Y^2 (three fractions of outliers k/n = 0,1; 0,2; 0,3 were considered).

The stages 4-8 were analogous as in the experiments *E1*. The results of the experiments *E2* are presented in tables 6.–10. (statistical characteristics of estimates and values of *EFG*) and on the graph 2 (illustration of *EFG* for k/n = 0,2).

Table 6. Statistical characteristics of DRM-estimates of parameters of model (5) for the experiments *E2*, k/n=0,2 and samples drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	2,001	0,008	0,008	2,982	0,057	0,057	1,220	0,312	0,439	0,205	0,018
200	2,001	0,005	0,005	2,989	0,037	0,037	1,207	0,210	0,327	0,205	0,017
300	2,000	0,004	0,004	2,992	0,029	0,029	1,202	0,167	0,285	0,205	0,017
400	2,000	0,003	0,003	2,993	0,025	0,025	1,201	0,144	0,265	0,205	0,017
500	2,000	0,003	0,003	2,995	0,023	0,023	1,197	0,130	0,251	0,205	0,017
1000	2,000	0,002	0,002	2,995	0,016	0,016	1,196	0,094	0,226	0,205	0,017

Source: author's calculations.

Table 7. Statistical characteristics of DRM-estimates of parameters of model (5) for the experiments *E2*, k/n=0,2 and samples drawn without replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	2,001	0,008	0,008	2,983	0,056	0,056	1,219	0,302	0,429	0,205	0,018
200	2,000	0,005	0,005	2,991	0,036	0,036	1,204	0,201	0,316	0,205	0,017
300	2,000	0,004	0,004	2,992	0,028	0,028	1,201	0,162	0,279	0,205	0,017
400	2,000	0,003	0,003	2,993	0,024	0,024	1,201	0,137	0,260	0,205	0,017
500	2,000	0,003	0,003	2,994	0,021	0,021	1,199	0,119	0,245	0,205	0,017
1000	2,000	0,002	0,002	2,996	0,013	0,013	1,195	0,078	0,216	0,205	0,017

Source: author's calculations.

Table 8. Statistical characteristics of LSM-estimates of parameters of model (5) for the experiments *E2*, *k/n*=0,2 and samples drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	2,001	0,002	0,002	2,989	0,020	0,020	1,029	0,160	0,167	0,015	0,016
200	2,001	0,001	0,002	2,988	0,013	0,014	1,033	0,109	0,117	0,015	0,015
300	2,001	0,001	0,001	2,987	0,011	0,011	1,035	0,087	0,097	0,015	0,015
400	2,001	0,001	0,001	2,987	0,009	0,010	1,035	0,075	0,085	0,015	0,015
500	2,001	0,001	0,001	2,987	0,008	0,009	1,035	0,067	0,078	0,015	0,015
1000	2,001	0,001	0,001	2,986	0,006	0,007	1,037	0,047	0,061	0,015	0,015

Source: author's calculations.

Table 9. Statistical characteristics of LSM-estimates of parameters of model (5) for the experiments *E2*, *k/n*=0,2 and samples drawn without replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
100	2,001	0,002	0,002	2,989	0,020	0,020	1,029	0,158	0,165	0,015	0,016
200	2,001	0,001	0,002	2,988	0,013	0,013	1,033	0,103	0,112	0,015	0,015
300	2,001	0,001	0,001	2,987	0,010	0,011	1,035	0,082	0,092	0,015	0,015
400	2,001	0,001	0,001	2,987	0,008	0,009	1,036	0,069	0,080	0,015	0,015
500	2,001	0,001	0,001	2,986	0,007	0,009	1,036	0,060	0,072	0,015	0,015
1000	2,001	0,000	0,001	2,986	0,004	0,006	1,037	0,037	0,053	0,015	0,015

Source: author's calculations.

Table 10. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E2* and *k/n* = 0,1; 02; 03

Sample size	Values of *EFG* for parameters																	
	k/n=0,1						*k/n*=0,2						*k/n*=0,3					
	A1		A2		A3		A1		A2		A3		A1		A2		A3	
	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM	DRM	LSM
100	1,8	-0,2	1,9	0,7	2,9	2,5	4,9	4,8	4,9	1,4	6,4	2,7	53,1	-0,1	6,6	-0,4	-6,8	0,4
200	4,3	14,3	4,0	11,9	5,4	12,6	7,5	12,5	8,4	9,5	9,7	10,7	12,6	7,1	11,4	8,4	10,5	9,6
300	9,3	25,0	7,9	14,5	9,2	14,7	7,1	12,5	6,9	12,5	6,7	12,4	7,1	12,3	6,6	11,0	7,0	11,0
400	16,7	21,4	17,1	23,3	16,7	21,8	15,8	33,3	13,9	20,6	11,8	19,2	9,2	25,0	9,4	16,8	9,7	16,4
500	27,3	50,0	24,5	29,9	24,1	29,2	21,4	50,0	20,5	22,2	19,1	22,9	17,4	33,3	16,4	19,8	15,5	20,5
1000	55,6	85,7	64,1	76,4	68,4	77,1	45,5	55,6	42,8	63,2	43,6	63,4	21,1	54,5	23,9	49,3	27,7	51,5

Source: author's calculations.

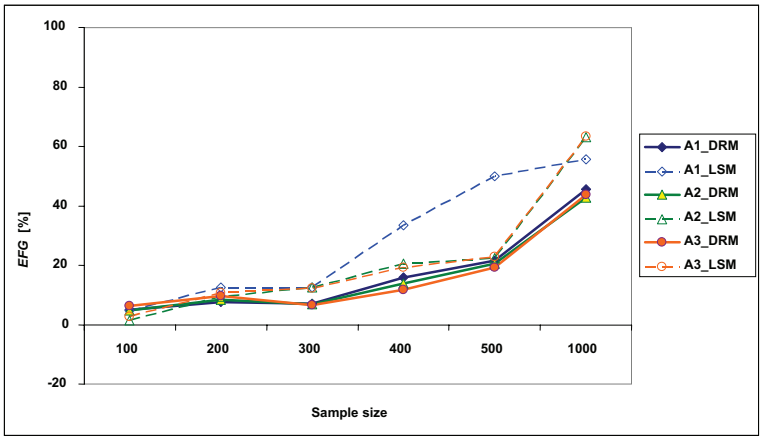

Fig. 2. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E2* and $k/n = 02$

Source: Table 10.

The last group of experiments *E3*, was conducted as follows:

1. Construction of the *N*-element ($N = 6000$) population, consisting of three strata: W1, W2 and W3, with the number of elements: 1200, 1800 and 3000, respectively. Population was described by the random variable (X_1, X_2, Y), where values of X_1, X_2 were generated independently from the chi-square distributions respectively: for W1: $X_1 \sim \chi^2_5$, $X_2 \sim \chi^2_{10}$, for W2: $X_1 \sim \chi^2_7$, $X_2 \sim \chi^2_{12}$ and for W3: $X_1 \sim \chi^2_9$, $X_2 \sim \chi^2_{14}$. Variable Y was constructed according to the formula (4).
2. Proportional drawing a *n*-element sample ($n = 50, 100, 200, 300, 400, 500, 1000, 1500$), proportional to the size of stratum (with replacement and without replacement for each stratum).
3. Input of the outliers into the samples by multiplication by 1,5 the first value of variable *Y* from each stratum.

The stages 4-8 were analogous as in experiments *E1*. The results of the experiments *E3* are presented in tables 11.–13. (statistical characteristics for DRM- and LSM- estimates for samples drawn with replacement and values of *EFG*) and on the graph 3 (illustration of *EFG*). Values of statistical characteristics for samples drawn without replacement were very close to obtained for samples drawn with replacement, so they were not presented in the paper.

Table 11. Statistical characteristics of DRM-estimates of parameters of model (5) for the experiments *E3* and sample drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
50	1,993	0,016	0,016	3,002	0,008	0,008	1,069	0,343	0,374	0,028	0,009
100	1,994	0,011	0,011	3,003	0,006	0,006	1,042	0,237	0,250	0,018	0,009
200	1,993	0,007	0,008	3,003	0,004	0,004	1,029	0,168	0,175	0,013	0,009
300	1,993	0,006	0,007	3,004	0,003	0,003	1,025	0,139	0,144	0,012	0,009
400	1,993	0,005	0,006	3,004	0,003	0,003	1,022	0,116	0,121	0,011	0,009
500	1,993	0,005	0,006	3,004	0,002	0,003	1,019	0,106	0,110	0,010	0,008
1000	1,993	0,003	0,005	3,004	0,002	0,002	1,017	0,075	0,077	0,009	0,008
1500	1,993	0,003	0,004	3,004	0,001	0,002	1,016	0,060	0,063	0,009	0,008

Source: author's calculations.

Table 12. Statistical characteristics of LSM-estimates of parameters of model (5) for the experiments *E3* and sample drawn with replacement

Sample size	PARAMETER									$\overline{RE}_s$	$\overline{RE}_{pop}$
	A1			A2			A3				
	MEAN	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*	*MEAN*	*V*	*RRMSE*		
50	1,998	0,010	0,009	3,003	0,005	0,005	0,989	0,224	0,222	0,008	0,009
100	1,998	0,006	0,006	3,003	0,003	0,003	0,991	0,152	0,151	0,008	0,009
200	1,999	0,005	0,004	3,003	0,002	0,002	0,987	0,108	0,108	0,008	0,008
300	1,998	0,004	0,004	3,003	0,002	0,002	0,988	0,087	0,086	0,008	0,008
400	1,998	0,003	0,003	3,003	0,002	0,002	0,987	0,075	0,075	0,008	0,008
500	1,998	0,003	0,003	3,003	0,001	0,002	0,987	0,067	0,068	0,008	0,008
1000	1,999	0,002	0,002	3,003	0,001	0,001	0,986	0,048	0,049	0,008	0,008
1500	1,998	0,002	0,002	3,003	0,001	0,001	0,986	0,039	0,040	0,008	0,008

Source: author's calculations.

Table 13. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E3*

Sample size	Values of *EFG* for parameters					
	A1		A2		A3	
	DRM	LSM	DRM	LSM	DRM	LSM
50	2,7	3,5	–0,6	–3,4	–0,1	–1,1
100	1,7	–1,3	3,3	0,8	2,7	–1,3
200	2,6	5,5	7,0	7,0	6,3	6,1
300	4,0	4,3	6,9	3,6	7,3	3,9
400	6,7	5,9	3,0	4,8	2,4	5,2
500	10,0	7,4	5,8	12,5	10,8	9,7
1000	18,2	16,7	20,8	28,6	21,8	23,2
1500	18,2	42,9	33,3	50,0	33,4	31,4

Source: author's calculations

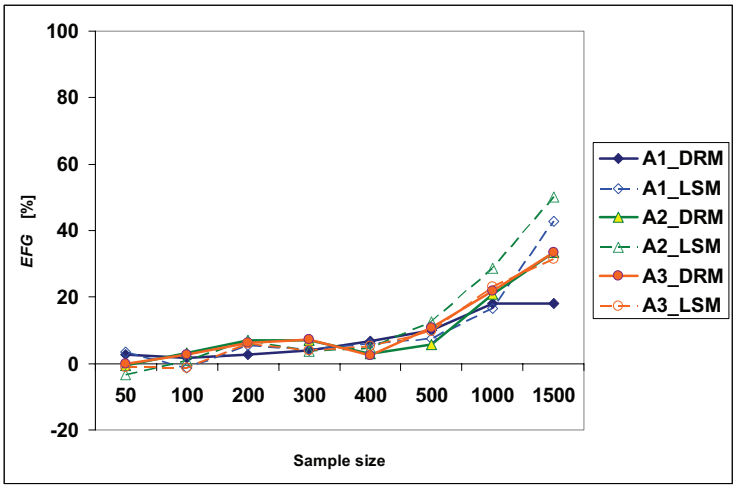

Fig. 3. Values of *EFG* for DRM- and LSM-estimates of parameters of model (5) for the experiments *E3*

Source: Table 13.

III. CONCLUSIONS FROM THE MONTE CARLO EXPERIMENTS

Results of all experiments presented in the paper led to the following conclusions:

- Mean of DRM- and LSM-estimates of parameters of model (5) were close to values of parameters of model (4) for all considered fraction of outliers in the sample.
- Values of the coefficient of variation and the relative root mean square error (*RRMSE*) for each parameter of model (5) decreased when the sample size increased for DRM- and LSM-estimates.
- In experiments E1 and E2 for DRM-estimates the coefficient of variation and the relative root mean square error (*RRMSE*) for each parameter of model (5) grew, when the fraction of outliers in the sample grew.
- Generally, values of the coefficient of variation and the relative root mean square error (*RRMSE*) for each parameter of model (5) for samples drawn without replacement were not higher than in case of samples drawn with replacement for both DRM- and LSM-estimates.
- Values of *EFG* increased for each parameter of model (5) when the sample size increased for both considered methods of estimation, but the growth for DRM was slower than for LSM.

- In experiments E1 and E2 in case of DRM-estimation values of *EFG* decreased, when the fraction of outliers in the sample grew.
- The mean of relative errors for sample ($\overline{RE}_s$) for DRM was higher than for LSM and than the mean relative errors for population ($\overline{RE}_{pop}$), what is connected with the unrobust, with respect to outliers, construction of measure of $\overline{RE}_s$.
- In most cases the value of the mean relative errors for population ($\overline{RE}_{pop}$) was comparable for both DRM- and LSM-estimation.

As a whole, regarding all the conducted experiments, the deepest regression method turned out to be robust with respect to the outliers for each considered sample size, fraction of outliers and sampling scheme. For data sets with outliers the DRM-estimation gave the outcomes comparable with results obtained with LSM after removing outliers from samples, but slightly worse. Sampling without replacement led to some growth of efficiency for estimator in comparison with sampling with replacement for both considered methods.

REFERENCES

Brandt S. (1999), *Analiza danych. Metody statystyczne i obliczeniowe*, PWN, Warszawa.
Rousseeuw P. J., Hubert M. (1999) *Regression Depth*, JASA 94, 388–402.
Van Aelst S., Rousseeuw P. J., Hubert M., Struyf A. (2002), *The Deepest Regression Method*, Journal of Multivariate Analysis, 81, 138–166.
Zasępa R. (1972), *Metoda reprezentacyjna*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.

Dorota Pruska

ODPORNOŚĆ METODY NAJGŁĘBSZEJ REGRESJI NA OBSERWACJE NIETYPOWE PRZY WYBRANYCH SCHEMATACH LOSOWANIA PRÓBY

Metoda najgłębszej regresji jest metodą szacowania parametrów funkcji regresji, odporną na występowanie obserwacji nietypowych w próbach wylosowanych z populacji.

W pracy przeprowadzono symulacyjną analizę odporności metody najgłębszej regresji na obserwacje nietypowe w przypadku wybranych schematów losowania prób. Na podstawie przeprowadzonych eksperymentów Monte Carlo wyznaczono charakterystyki rozkładu ocen parametrów modelu regresji liniowej otrzymanych metodą najgłębszej regresji na podstawie prób zawierających obserwacje nietypowe. Uzyskane wyniki porównano z wynikami analogicznych eksperymentów, w których estymacji parametrów dokonano metodą najmniejszych kwadratów po uprzednim usunięciu z prób obserwacji nietypowych. Metoda najgłębszej regresji okazała się odporna na obserwacje nietypowe we wszystkich rozważanych przypadkach, a jej wyniki porównywalne z uzyskanymi metodą najmniejszych kwadratów po usunięciu z prób obserwacji nietypowych, choć nieco gorsze.